NAA10 polyadenylation signal variants cause syndromic microphthalmia

Jennifer J. Johnston[1*], Kathleen A. Williamson[2], Christopher M. Chou[3^], Julie C. Sapp[1], Morad Ansari[2,4], Heather M. Chapman[5], David N. Cooper[6], Tabib Dabir[7], Jeffrey N. Dudley[1], Richard J. Holt[8], Nicola K. Ragge[8,9], Alejandro A. Schäffer[10,11], Shurjo K. Sen[12], Anne M. Slavotinek[13], David R. FitzPatrick[2], Tom M. Glaser[5], Fiona Stewart[7], Graeme C.M. Black[14,15], Leslie G. Biesecker[1**]

1. National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
2. MRC Human Genetics Unit, IGMM, University of Edinburgh, Edinburgh EH4 2XU UK
3. Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, USA
4. DNA Diagnostic Laboratory, South East Scotland Regional Genetics Services, Western General Hospital, Edinburgh EH4 2XU UK
5. Department of Cell Biology and Human Anatomy, U.C. Davis, Davis, CA, USA
6. Institute of Medical Genetics, Cardiff University, Heath Park, Cardiff CF14 4XN, UK
7. Northern Ireland Regional Genetics Service (NIRGS), Belfast City Hospital, Belfast, Northern Ireland, UK
8. Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, UK
9. West Midlands Regional Clinical Genetics Service and Birmingham Health Partners, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK

10. Computational Biology Branch, National Center for Biotechnology Information, NIH, Bethesda, Maryland, USA
11. Cancer Data Science Laboratory, National Cancer Institute, NIH, Bethesda, MD, USA
12. Leidos Biomedical Research, Inc, Basic Science Program, Cancer & Inflammation Program, Frederick National Laboratory for Cancer Research, Bethesda, Maryland 20892, USA
13. Department of Pediatrics and Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA
14. Manchester Centre for Genomic Medicine, Manchester Academic Health Sciences Centre, Manchester University NHS Foundation Trust, St Mary's Hospital, Manchester, UK
15. Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

Corresponding Authors

*Jennifer J. Johnston, PhD

50 South Drive Room 5139, Bethesda, MD 20892

Tel: 301-594-3981 Fax: 301-402-2170

Email: jjohnsto@mail.nih.gov

**Leslie G. Biesecker, MD

50 South Drive Room 5140, Bethesda, MD 20892

Tel: 301-402-2041 Fax: 301-402-2170

Email: lesb@mail.nih.gov

^Current affiliation, Kaiser-Permanente Health System, Sacramento, CA, USA

Word Count: 4,000

CONFLICT OF INTEREST PAGE

LGB receives royalties from Genentech Corp, is an advisor to the Illumina Corp, received honoraria from Wiley-Blackwell and receives honoraria from Cold Spring Harbor Press. LGB, JJJ, JND, and JCS received support from the Intramural Research Program of the National Human Genome Research Institute grants HG200328 11 and HG200388 03. RJH and NKR were supported by grants from Baillie Gifford, Visually Impaired Children Taking Action (VICTA) (http://www.victa.org.uk/), and Microphthalmia, Anophthalmia, Coloboma Support (MACS) (www.macs.org.uk). AAS received support from the Intramural Research Program of the National Library of Medicine. DNC received funding from Qiagen Inc. through a License Agreement with Cardiff University. AMS received support from the extramural research program of the National Human Genome Research Institute grant 5U01HG009599 02, and received honoraria from Wiley-Blackwell, Inc., Oxford University Press and UptoDate, Inc. TMG and HMC received support from the National Institutes of Health grant EY19497 and HMC received support from the Hartwell Foundation. DRF and KAW were funded via the MRC University Unit award to the University of Edinburgh for the MRC Human Genetics Unit. MA was partially funded by the MRC IGMM Translational Science Initiative.

Abstract

Background: A single variant in *NAA10* (c.471+2T>A), the gene encoding N-acetyltransferase 10, has been associated with Lenz microphthalmia syndrome. In this study, we aimed to identify causative variants in families with syndromic X-linked microphthalmia.

Methods: Three families, including 15 affected individuals with syndromic X-linked microphthalmia, underwent analyses including linkage analysis, exome sequencing and targeted gene sequencing. The consequences of two identified variants in *NAA10* were evaluated using quantitative PCR and RNAseq.

Results: Genetic linkage analysis in family 1 supported a candidate region on Xq27-28, which included *NAA10*. Exome sequencing identified a hemizygous *NAA10* polyadenylation signal (PAS) variant, chrX:153,195,397T>C, c.*43A>G, which segregated with the disease. Targeted sequencing of affected males from families 2 and 3 identified distinct *NAA10* PAS variants, chrX:g.153,195,401T>C, c.*39A>G and chrX:g.153,195,400T>C, c.*40A>G. All three variants were absent from gnomAD. Quantitative PCR and RNAseq showed reduced *NAA10* mRNA levels and abnormal 3' UTRs in affected individuals. Targetted sequencing of *NAA10* in 376 additional affected individuals failed to identify variants in the PAS.

Conclusion: These data show that PAS variants are the most common variant type in *NAA10*-associated syndromic microphthalmia, suggesting reduced RNA is the molecular mechanism by which these alterations cause microphthalmia/anophthalmia. We reviewed recognized variants in PAS associated with Mendelian disorders and identified only 23 others, indicating that *NAA10* harbors more than 10% of all known PAS variants. We hypothesize that PAS in other genes harbor unrecognized pathogenic variants associated with Mendelian disorders. The systematic

interrogation of PAS could improve genetic testing yields.

Introduction

Congenital anophthalmia and microphthalmia result from failure of early eye development. One-third of affected individuals are syndromic[1],[2] and it is genetically heterogeneous.[3] X-linked syndromic anophthalmia and microphthalmia have been shown to result from pathogenic variants in four genes: *BCOR*[4] (MIM:300485, MCOPS2), *HCCS*[5] (MIM:300056, MCOPS7), *HMGB3*[6] (MIM:300193, MCOPS13), and *NAA10*[7] (MIM:309800, MCOPS1). A fifth locus, MCOPS4 (MIM:301590, also designated as ANOP1), with linkage to Xq27-q28, was specified without identification of an associated gene. *HMGB3* and *NAA10* reside on Xq27-Xq28. A single *HGMB3* variant has been reported in one family to cause syndromic colobomatous microphthalmia.[6] Forrester et al. described a family segregating what they termed Lenz microphthalmia syndrome (LMS) and linkage to Xq27-q28.[8] Affected individuals had severe microphthalmia, renal anomalies, high-arched palate, cutaneous syndactyly of the hands, and severe intellectual disability (ID). A c.471+2T>A splice variant in *NAA10* was identified as causative, the only study to date associating *NAA10* variants with syndromic microphthalmia.[7]

The MCOPS4 locus was originally defined based on a family reported by Graham et al.[9] Their manifestations included microphthalmia/anophthalmia and ID in multiple affected males. Here, we re-analyze this family using genetic linkage analysis and exome sequencing, verified our findings in two additional families, and functionally characterized the variants.

Materials and Methods

This study was approved by Institutional Review Boards at the National Institutes of Health, the UK Multicentre Research Ethics Committee (MREC), the University of Michigan, the University

of California, Davis, and the Huntingdon Ethics Committee, UK. All DNA analyses were performed using standard techniques.

Linkage and haplotype analyses

Resampling and repeat genetic linkage analyses were performed in family 1 (Figure 1A) using 30 microsatellite markers on the X chromosome. Initial analyses of 22 markers spanning the X chromosome were followed by fine mapping with an additional eight markers from the region near Xqtel. LOD scores were computed using FASTLINK using equal marker allele frequencies.[10], [11], [12] For multi-marker analyses, inter-marker recombination fractions were derived from the Rutgers map.[13] The disease locus was modeled as X-linked recessive. Analyses were performed under two additional possible scenarios, first that the individual I-2 was gonadal mosaic and second that penetrance was less than 100%. Haplotype analysis was performed assuming a model of X-linked recessive inheritance.

Sequence analyses

The coding region and consensus splice sites of two candidate genes, *NAA10* and *HMGB3,* were analyzed in family 1 using Sanger sequencing. Exome sequencing was performed on individual IV-3 and his parents at the NIH Intramural Sequencing Center (NISC), as described.[14] Identified variants in the linkage region were filtered for quality, absence in ExAC males and absence in 473 males from an in-house dataset.[15],[16] Sanger sequence analyses were performed in family 1 to confirm segregation. Targeted Sanger sequencing of *NAA10* was performed in the proband of family 2 (Figure 1B). Independently, sequence analysis of a panel of 41 genes causally associated with severe eye malformations was carried out on the proband in family 3 (Figure 1C) using a custom-designed Agilent SureSelect panel (Agilent Technologies,

Inc., Santa Clara, CA) and paired-end DNA sequencing using an Illumina MiSeq platform (Illumina, San Diego, CA), as per the manufacturer's instructions. Targeted sequencing of the *NAA10* polyadenylation signal (PAS) was performed in 250 unrelated microphthalmia/anophthalmia/coloboma patients selected without sex bias from the MRC Human Genetics Unit Eye Malformation cohort. A further 126 males with developmental eye disorders underwent targeted sequencing of the *NAA10* PAS. Forty-four of these individuals with bilateral developmental eye disorders were analysed for variants in the coding region and consensus splice sites of *NAA10*.

X-inactivation analyses

X-inactivation studies were performed on genomic DNA from 11 carrier females (family 1 individuals I-2, II-2, II-6, II-17, III-15; family 2 individuals III-2, III-3, IV-1, IV-5, V-8; family 3 individual I-2) using the human androgen receptor (HUMARA) assay as described in Allen et al[17] with minor modifications. Briefly, genomic DNA was digested with *Rsa1* either alone or with the addition of *HpaII* (New England Biolabs, Ipswich, MA). Digested DNA was PCR amplified using primers HUMAR_F: /56-FAM/TCCAGAATCTGTTCCAGAGCGTGC and Humara_R: GCTGTGAAGGTTGCTGTTCCTCAT (Integrated DNA Technologies, Inc. Coralville, Iowa). Products were run on an 3130xl genetic analyzer (ThermoFisher Scientific, Waltham, MA) and analyzed using GeneMapper v4.0 (ThermoFisher Scientific, Waltham, MA).

RNA Analyses

Total RNA was isolated from whole blood from affected and carrier individuals in family 1 and from unrelated male and female controls using PAXgene Blood RNA tubes (BD Biosciences, San Jose, CA). Total RNA was isolated from lymphoblasts from one affected and one unaffected

individual in family 2 using the RNeasy Mini kit (Qiagen, Germantown, MD). RNA samples were not available from family 3. RNA-Seq libraries were constructed from 0.5-1 µg total RNA after rRNA depletion using Ribo-Zero Globin (Illumina, San Diego, CA). The Illumina TruSeq Stranded Total RNA Kit was used according to manufacturer's instructions. The cDNA inserts were ~200 bp after chemical shearing. PCR amplification was performed using 10 cycles. Unique barcode adapters were applied to each library. Libraries were pooled in equimolar ratio for sequencing. The pooled libraries were sequenced on multiple lanes of a HiSeq 2500 (Illumina, San Diego, CA) using version 4 chemistry to achieve a minimum of 69 million 125 base read pairs. The data were processed using RTA version 1.18.64 and CASAVA 1.8.2. RNA-Seq reads were aligned with the STAR aligner,[18] using the hg19 genome assembly and the Ensembl transcript database (release 74). The STAR BAM output file was converted to WIG format using the bam2wig.py script from the RSeqC suite,[19] and these WIG files were further converted to bigwig format using the wigToBigWig tool from the UCSC Genome Browser's binary utilities directory.

For quantitative PCR (qPCR), 400 ng total RNA was converted to cDNA using the High Capacity RNA-to-cDNA kit (Applied Biosystems, Beverly, MA). TaqMan assays were performed using 10 ng cDNA, gene specific TaqMan probes (*NAA10*: HS01125831_g1 and HA00185854_m1; beta-Actin: Hs99999903_m1 (Thermo Fisher Scientific, Waltham, MA)) and TaqMan Gene Expression Master Mix (Applied Biosytems, Beverly, MA) on the OneStepPlus Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA). Samples were run in triplicate and relative expression levels in affected and carrier individuals as compared to control individuals were calculated using the delta-delta C_t method.

Results

Clinical evaluations

Family 1

Family 1 (Figure 1A) was originally described by Graham et al. in support of a fifth microphthalmia/anophthalmia locus MCOPS4. The family was reported to include seven affected males with microphthalmia/anophthalmia and ID.[9] One male, II-18, was reported to have normal eyes with cleft soft palate and a single ear tag and was considered unaffected with respect to syndromic anophthalmia/microphthalmia. An additional affected male, IV-3, with a milder phenotype as compared to his affected uncles, was born after publication of the Graham et al paper.[9] Briefly, this male infant was born following a pregnancy in which scans were assessed to be normal. However, at birth he was noted to have a large open neural tube defect (NTD), right sided anophthalmia, and bilateral 2-3 cutaneous syndactyly of his toes. When his teeth erupted they were irregular. He was sociable and did not show signs of the severe developmental delay or behavioral problems, most notably self-mutilation, seen in other affected males in the family.

Family 2

Family 2 (Figure 1B) was originally described by Slavotinek et al.[20] as a four-generation family with X-linked anophthalmia consistent with linkage to Xq27. The intelligence of two living males in family 2 was apparently normal. Carrier females III-2 and IV-1 had unremarkable eye examinations, individual IV-1 was diagnosed with dyslexia but her intelligence was apparently normal.

Family 3

The male proband (individual II-2, Figure 1C) was the second child of a non-consanguineous couple, referred for diagnostic screening of genes associated with severe eye malformations. Family history was remarkable for a maternal uncle with anencephaly. The proband presented with spina bifida at 20 weeks gestation. He was a term delivery with birth weight 3.2 Kg. At birth, he was noted to have widely spaced eyes, unilateral phthisis bulbi, downturned corners of the mouth, small penis, and small feet with upturned nails. At eight months his length, weight, and head circumference were recorded at 9th centile. He had a repaired myelomeningocele, hydrocephalus with a ventriculoperitoneal shunt, and left grade 2 vesico-ureteric reflux with a normal renal ultrasound. MRI studies showed ventriculomegaly with a Chiari type 2 malformation. Developmental milestones were reported to be normal at 8 months. Array CGH was reported as normal.

Molecular Data

Family 1

LOD score analyses reported in the original paper showed a peak logarithm of odds (LOD) score of 1.9, between the disease and markers on Xq27-q28, below the 2.0 threshold recommended to declare X-chromosome linkage.[21] The evidence of linkage was weak unless one assumed non-penetrance, pleiotropy, variable expressivity, or gonadal mosaicism as the mutant haplotype was shared by II-18 and his affected relatives.

Repeat multi-marker analysis using four informative markers DXS8091, DXS1193, DXS8086, and DXS1073 achieved a peak LOD score of 1.43 with the disease locus placed on top of DXS1073 and assuming full penetrance. The peak multi-marker score was 2.19 with 75% penetrance, 2.46 with 50% penetrance, and 3.31 assuming gonadal mosaicism. Haplotype analysis was

performed assuming a model of X-linked recessive inheritance and showed a recombination event in affected individual III-2 between markers DXS8043 and DXS8086 restricting the possible linkage region telomeric to position ChrX(GRCh37):g.144,028,513.

Both *HMGB3* and *NAA10* lie within this linkage region. The coding regions and consensus splice sites of *HMGB3* and *NAA10* were interrogated without identification of a pathogenic variant. Trio exome sequencing was performed on individual IV-3 and his parents and variants in the linked region were filtered for absence in male controls. A single variant was identified in the 3'UTR of *NAA10*, chrX:153,195,397T>C, c.*43A>G (reference cDNA NM_003491.3), altering the consensus PAS from AATAAA to AATAGA. The *NAA10* variant was verified by Sanger sequencing and shown to be present in all affected individuals available for analysis, the matriarch (I-2), as well as in individual II-18, the male previously thought to be unaffected. Querying the gnomAD database for variants in the consensus PAS for *NAA10* did not identify any variants in over 170,000 alleles.

Family 2

A reexamination of the linkage data showed that family 1 and family 2 had overlapping regions near Xqtel consistent with linkage. Based on the findings in family 1, targeted Sanger sequencing of *NAA10* was performed in the proband of family 2. A second distinct variant in the the 3'UTR of *NAA10*, chrX:g.153,195,401T>C, c.*39A>G, was identified, which altered the consensus PAS from AATAAA to GATAAA. The variant was verified by Sanger sequencing in carrier females and was not present in two unaffected males.

Family 3

Targeted sequencing of 41 genes causally associated with severe eye malformations in the proband from family 3 identified a third variant in *NAA10*, chrX:g.153,195,400T>C , c.*40A>G, that altered the consensus PAS from AATAAA to AGTAAA. The *NAA10* variant in family 3 was confirmed by Sanger sequencing. DNA from the uncle with anencephaly was not available for testing.

***NAA10* Sequence Analysis in Microphthalmia/Anophthalmia/Coloboma Cohorts**

Targeted sequencing of the *NAA10* PAS in 250 unrelated microphthalmia/anophthalmia/coloboma patients selected without sex bias from the MRC Human Genetics Unit Eye Malformation cohort did not identify additional variants. No variants affecting the *NAA10* PAS were identified in a further 126 males with developmental eye disorders. In addition, no pathogenic variants in the coding region and consensus splice sites of *NAA10* were identified in 44 of these individuals who presented with bilateral developmental eye disorders.

X-Inactivation

Four of 11 carrier females showed greater than 90% skewing of X-inactivation, individual II-2 from family 1 and individuals III-2, IV-1 and IV-5 from family 2. However, females did not show consistent skewing of X-inactivation.

RNA analyses

The consequences of the PAS variants for *NAA10* mRNA expression level and structure were investigated using qPCR and RNAseq in families 1 and 2. qPCR showed carrier females from family 1 to have similar levels of *NAA10* RNA as compared to control individuals. In contrast,

affected individuals from families 1 and 2 demonstrated a decrease in the quantity of *NAA10* RNA of approximately 50% when compared to controls (Figure 2).

Transcript structure was investigated by mapping RNAseq reads to the UCSC Genome Browser. RNAseq data from unaffected (family 2, individual V-9) and control individuals indicated the normal 3'UTR of *NAA10* ended at the approximate position of the predicted polyadenylation site (Figure 3G-I). In contrast, for affected individuals with the PAS variants (family 1, individuals III-2, III-5, IV-3 and family 2, individual IV-6), the read depth did not decrease as expected at the polyadenylation site in the 3'UTR, but instead declined approximately 600 bp further 3' at a second polyadenylation site predicted by bioinformatic analysis of genomic sequence (Figure 3A-C, F).[22] A similar result was observed in carrier females (family 1, individuals II-2 and III-15) (Figure 3D-E).

Discussion

Six protein complexes, NatA through NatF, carry out N^{α} acetylation of proteins in the cell.[23],[24] NAA10 is the primary subunit of the NatA complex and an auxiliary subunit of the NatE complex.[24],[25] The acetylation targets of the six complexes are believed to be mostly distinct and dependent on the amino acids that follow the methionine in position 1.[24] Additionally, data from a knock-out mouse model suggests that NAA10 plays a role in the regulation of methylation through direct DNA binding.[26] Prior to the present publication, the mutational spectrum in *NAA10* comprised nine missense alterations and the single splice site variant identified previously in a single family with what those authors called LMS. The first reported variant in *NAA10*, c.109T>C; p.(Ser37Pro), was identified in two families segregating a sex-linked recessive male lethal syndrome with an aged appearance and cardiac

arrhythmias.[27] Eight additional missense variants have since been reported in heterozygous females or hemizygous males with developmental delay and/or ID with or without cardiac involvement and without anophthalmia.[28],[29],[30],[31],[32] The majority of these variants were identified as *de novo* alterations.[28],[30],[31],[32] Functional studies of the missense alterations have demonstrated a variable effect on the ability of NAA10 to acetylate substrates and the clinical phenotype may correlate with the level of acetylation dysfunction and affected substrates. Alternatively, clinically relevant predicted missense variants in *NAA10* have been shown to disrupt DNA binding[26] and it is likely that different missense alterations will variably contribute to this phenomenon.

The polyA_DB database contains human mRNA polyadenylation sites based on EST/cDNA evidence[33] and predicts four sites in the 3' region of *NAA10*. Based on the *NAA10* Poly(A) sequence data in the UCSC Genome Browser, the PAS at c.*39 – c.*44 is the most commonly used signal with polyadenylation occurring between 12-19 nucleotides 3' of that hexamer. Three distinct nucleotides within this six-nucleotide signal are mutated in the families described here. Previous in vitro studies have demonstrated that these changes in the PAS hexamer disrupt both cleavage and polyadenylation (Figure 4).[34] Furthermore, these variants are analogous to variants associated with abnormal phenotypes in other genes. The *NAA10* 3'UTR variant identified in family 1, c.*43A>G, alters the AATAAA consensus to AATAGA, similar to the c.*112A>G (reference cDNA NM_000518.4) pathogenic variant in the beta globin (*HBB)* gene (HGMD CR900266) (See Table 1).[35] The 3'UTR variant of *NAA10* identified in family 2, c.*39A>G, alters the AATAAA consensus to GATAAA, similar to another reported pathogenic variant in the *HBB* gene, (c.*108A>G, HGMD CR127145).[36] The *NAA10* 3'UTR variant

identified in family 3 , c.*40A>G, alters the AATAAA consensus to AGTAAA, similar to the variant in the *ARSA* gene (HGMD CR890137) reported by Gieselmann et al[37] and a variant in the *SLC6A4* gene (HGMD CR102248) reported by Gyawali et al[38] although it is important to note that the wild-type PAS consensus sequence in both *ARSA* and *SLC6A4* genes is AATAAC, (See Table 1). That the *NAA10* variant disrupts a consensus PAS rather than a non-canonical PAS could be consistent with the observation that the *NAA10* variants are associated with a Mendelian disorder whereas the *ARSA* and *SLC6A4* variants are associated with lower penetrance traits.

All three families presented here manifested microphthalmia/anophthalmia with variable additional features. In family 1, individuals had variable expressivity ranging from syndromic microphthalmia/anophthalmia with severe ID to isolated cleft palate and an ear tag. A single individual, IV-3, had an open NTD. Affected individuals in family 2 had anophthalmia and additional findings including minor skeletal anomalies and attention deficit disorder, but notably all affected males were reported to have normal intelligence. The proband in family 3 had syndromic microphthalmia/anophthalmia with complex features including spina bifida. By report, a maternal uncle (deceased) to the proband had anencephaly, however, his genotype was unknown. Although the numbers are small, two out of 15 individuals in this study had NTDs (three if the uncle in family 2 is included). Additional families will need to be studied to determine if NTDs are a feature of *NAA10* PAS variants. The mouse *NAA10* knock-out model had variable expressivity ranging from embryonic lethality to normal body size and Lee et al. suggested this variable expressivity may result from the role of NAA10 in global methylation.[26] The role of NAA10 in global methylation may contribute to the observed

variable expressivity in these families as well. Specifically, the observed phenotype may result from the regions of the genome experiencing defective methylation in any one individual. As methylation is important in the etiology of NTDs it is interesting to speculate that the role of NAA10 in global methylation may account for the NTDs seen in these families. Future studies will be needed to determine if methylation or other factors play a role in the variable expressivity seen with these variants.

Cleavage and polyadenylation signals reside in the 3' UTRs of mRNAs and include a number of key *cis* elements including the AAUAAA PAS which typically resides 10-30 nucleotides upstream of the cleavage site (reviewed in [39],[40]). Ninety percent of identified PAS conform to the hexanucleotide sequence AAUAAA and most of the remaining 10% differ by only a single substitution.[39] The AAUAAA hexamer is critical both for cleavage and poly(A) addition,[41] and the effects of point mutations in this hexamer have been determined.[34] Disease-associated variants in PAS are rare and consequently we were surprised to identify three such variants in this rare disorder. To identify all known consensus PAS variants, we performed a directed search of the Human Gene Mutation Database (HGMD).[42] We identified 31 reported PAS variants in 19 genes (Table 1), 23 associated with Mendelian disorders and eight associated with a functional polymorphism or susceptibility. Overall, PAS variants account for 23 out of the ~195,000 DM variants in HGMD, or 0.012% of the total. As well as being individually uncommon, such variants have a highly non-random distribution across genes. Of the 23 variants associated with Mendelian disorders, nine are in *HBB*, four are in *HBA2*, and two are in *FOXP3*, the remainder being singletons. Some of this distribution can be explained by the depth and detail to which the various genes have been studied, *HBB* being one

of the best studied of all human genes. To allow for this, the proportion of polyadenylation variants to all known pathogenic variants in a given gene can be tabulated. When expressed in this way, a very different pattern emerges. *NAA10* has the highest proportion of PAS variants, at 25%. The next highest cluster of genes includes *IGF1*, *NAT1*, *BMP1*, and *SLC6A4*, each of which has a single PAS variant, but a relatively small overall number of variants, yielding a high proportion. *HBB* and *HBA2* lie in the middle of the range at about 1% of variants, reflecting the large number of variants in these well-studied genes. Most of the clinical phenotypes associated with PAS variants are typical of those seen for loss of function variants in the same gene although they may be found with milder forms of the phenotypes – for example, in *HBB,[36],[43]* the phenotypes of PAS variants are described as either typical or mild thalassemia. The contribution to milder phenotypes has two implications, first in disorders where loss of gene function would be lethal, PAS variants may allow sufficient function for viability. Second, for disorders where the phenotype is due to complete loss of function, PAS variants may be missed if they lead to a milder condition.[36] We conclude from these data that the phenotypic consequence of pathogenic PAS variants is likely due to loss or partial loss of function, consistent with the experimental observations of these variants in other genes and the qPCR data we show here. Specifically, for *NAA10*, this suggests that partial loss of protein function due to reduced mRNA levels leads to a distinct phenotype, as compared to missense variants where the functional deficiency may only affect specific acetylation or DNA binding functions. The previously reported *NAA10* splice site variant[7] also showed reduced mRNA/protein supporting reduced function as causal for the microphthalmia phenotype. Alternatively, sequences within the elongated 3' end, or within the retained intron 7 for the

splice variant, may have a novel functional effect. Possible novel effects include: altered RNA or protein localization[44], altered 3'UTR-directed protein-protein interactions[45], or acquisition of novel regulatory functions. However, since the retained sequences are distinct between the splice site variant and the PAS variants, a novel property of the mRNA is unlikely to be the mechanism accounting for the overlapping feature of microphthalmia. Additionally, the absence of a phenotype in carrier females without skewed X-inactivation suggests that any effect would need to fit a recessive model of inheritance. The general dearth of recognized PAS pathogenic variants may be in part due to the inherent difficulty in identifying PAS. In addition, many such variants may not be included in next generation sequence data due to the limitations of exome sequencing kits and the predominant use of exome over genome sequencing.

The assessment of pathogenicity for non-coding variants in Mendelian disorders is challenging. The ACMG/AMP [46] framework includes few criteria that are relevant to 3' UTR variants, reflecting a focus on high penetrance coding variants. For the variants we identified here, we invoked criteria PS3 (well-established in vitro or in vivo functional studies), PM1 (located in a mutational hot spot and/or critical and well-established functional domain), PM2 (absent from controls), and PP1 (co-segregation with disease in multiple affected family members). As qPCR is not "a well-established functional assay" because qPCR for *NAA10* is not established in the diagnostic realm, the PS3 criteria can be downgraded to a moderate level of evidence (PS3_Mod). Additionally, our segregation data are arguably less than robust considering that we have invoked variable expressivity to explain the occurrence of the variant in II-18 in family 1 and thus PP1 might arguably be dropped. That would leave PS3_Mod, PM1,

and PM2, which combine to yield a likely pathogenic assessment. Based on these pieces of evidence, we conclude that these three PAS variants are likely pathogenic.

These data have important implications for the overall process of identifying pathogenic variation. We hypothesize that PAS variants are candidate pathogenic variants for many Mendelian disorders and may be a component of the mutational spectrum of patients who are currently 'mutation-negative'. A full assessment of this hypothesis will first require the continued development of genome (as opposed to exome) population sequence reference databases, allowing a greater understanding of the evolutionary constraints operating at 3' UTR loci. Second, robust methods to identify PAS for all genes will need to be developed. While trivial for many genes, this is challenging for others as some use multiple polyadenylation sites and often the PAS sequence does not conform to the consensus, as demonstrated for a number of the genes that we reviewed. Third, to the extent that exome analysis continues to outpace genome analysis, it will be necessary to target these sequences in exome capture reagents. Fourth, testing laboratories will need to direct their attention to detecting and interpreting such variants. Finally, criteria for pathogenicity of non-coding variants will need to be updated. We believe that the accomplishment of these improvements in variant detection, interrogation, and interpretation will allow the findings based on the three variants reported here as a cause of a rare disorder to yield benefits for the diagnosis of patients with many other genetic disorders and contribute to the overall improvement in clinical genomic sequencing diagnostic yield.

Contributions

All authors fulfil the criteria for authorship. AAS performed and interpreted linkage analysis of family 1. JJJ permormed haplotype analysis, interpreted NGS data and performed Sanger confirmation in family 1. JCS, FS and LGB provided clinical management for family 1. HMC performed sequence analysis of *NAA10* in family 2. CMC and AMS provided clinical management for family 2. JND performed and interpreted qPCR experiments in families 1 and 2. SKS analyzed RNA-seq data for families 1 and 2. KAW and MA interpreted NGS data in family 3 and analyzed sequence data in a cohort of individuals with anophthalmia/microphthalmia for variants in *NAA10*. TD provided clinical management for family 3. RJH and NKR performed and analyzed sequencing of *NAA10* in a cohort of individuals with anophthalmia/microphthalmia/coloboma. DNC provided an analysis of polyadenylation signal variants within the Human Gene Mutation Database. JJJ and LGB drafted the manuscript and all authors contributed to the manuscript. All authors read and approved the final manuscript. LGB (family 1), GCMB (family 1), TMG (family 2) and DRF (family 3) planned the study. JJJ submitted the study.

Funding

This work was supported by funding from the Intramural Research Program of the National Human Genome Research Institute (LGB, JJJ, JCS, grants HG200328 11 and HG200388 03). This work was supported in part by funding from the Intramural Research Program of the National Library of Medicine (AAS, grant LM00097), the extramural research program of the National Human Genome Research Institute (AMS, grant 5U01HG009599 02), the National Institutes of Health (TMG, HMC, grant EY19497), the Hartwell Foundation (HMC), the MRC University Unit award to the University of Edinburgh for the MRC Human Genetics Unit (DRF, KAW), and the MRC IGMM Translational Science Initiative (MA). This work was supported in part by grants from Baillie Gifford, Visually Impaired Children Taking Action (VICTA) (http://www.victa.org.uk/), and Microphthalmia, Anophthalmia, Coloboma Support (MACS) (www.macs.org.uk).

Competing Interests

LGB receives royalties from Genentech Corp, is an advisor to the Illumina Corp, received honoraria from Wiley-Blackwell, and receives honoraria from Cold Spring Harbor Press. DNC is in receipt of funding from Qiagen Inc through a License Agreement with Cardiff University. AMS receives honoraria from Wiley-Blackwell, Inc., Oxford University Press and UptoDate, Inc.

Acknowledgments

We thank the families for their participation in this study and the NIH Intramural Sequencing Center for performing the exome sequencing on family 1 and the RNAseq data generation. We thank MaryPat Jones and Ursula Harper for generating STRP genotypes for this project. We are grateful to Dorine Bax for assisting with the coordination of the project. We thank Dr. Michael Sheets and Dr. Marvin Wickens for thoughtful review of the manuscript and permission to use Figure 4 reproduced from Sheets et al. 1990. The image was used by permission of Oxford University Press. We thank Catherine Driscoll for help in editing the manuscript.

WEB RESOURCES

Online Mendelian Inheritance in Man (OMIM): http://www.omim.org/

Human Gene Mutation Database (HGMD): http://www.hgmd.cf.ac.uk

University of California Santa Cruz genome browser (UCSC): https://genome.ucsc.edu/

REFERENCES

1. Tucker S, Jones B, Collin R. Systemic anomalies in 77 patients with congenital anophthalmos or microphthalmos. *Eye (Lond)* 1996;**10 (Pt 3)**:310-4 doi: 10.1038/eye.1996.65published Online First: 1996/01/01].
2. Verma AS, Fitzpatrick DR. Anophthalmia and microphthalmia. *Orphanet J Rare Dis* 2007;**2**:47 doi: 10.1186/1750-1172-2-47published Online First: 2007/11/28].
3. Gerth-Kahlert C, Williamson K, Ansari M, Rainger JK, Hingst V, Zimmermann T, Tech S, Guthoff RF, van Heyningen V, Fitzpatrick DR. Clinical and mutation analysis of 51 probands with anophthalmia and/or severe microphthalmia from a single center. *Mol Genet Genomic Med* 2013;**1**:15-31 doi: 10.1002/mgg3.2published Online First: 2014/02/06].
4. Ng D, Thakker N, Corcoran CM, Donnai D, Perveen R, Schneider A, Hadley DW, Tifft C, Zhang L, Wilkie AO, van der Smagt JJ, Gorlin RJ, Burgess SM, Bardwell VJ, Black GC, Biesecker LG. Oculofaciocardiodental and Lenz microphthalmia syndromes result from distinct classes of mutations in *BCOR*. *Nat Genet* 2004;**36**:411-6 doi: 10.1038/ng1321published Online First: 2004/03/09].
5. Wimplinger I, Morleo M, Rosenberger G, Iaconis D, Orth U, Meinecke P, Lerer I, Ballabio A, Gal A, Franco B, Kutsche K. Mutations of the mitochondrial holocytochrome c-type synthase in X-linked dominant microphthalmia with linear skin defects syndrome. *Am J Hum Genet* 2006;**79**:878-89 doi: 10.1086/508474published Online First: 2006/10/13].
6. Scott AF, Mohr DW, Kasch LM, Barton JA, Pittiglio R, Ingersoll R, Craig B, Marosy BA, Doheny KF, Bromley WC, Roderick TH, Chassaing N, Calvas P, Prabhu SS, Jabs EW. Identification of an *HMGB3* frameshift mutation in a family with an X-linked colobomatous microphthalmia syndrome using whole-genome and X-exome sequencing. *JAMA Ophthalmol* 2014;**132**:1215-20 doi: 10.1001/jamaophthalmol.2014.1731published Online First: 2014/07/06].
7. Esmailpour T, Riazifar H, Liu L, Donkervoort S, Huang VH, Madaan S, Shoucri BM, Busch A, Wu J, Towbin A, Chadwick RB, Sequeira A, Vawter MP, Sun G, Johnston JJ, Biesecker LG, Kawaguchi R, Sun H, Kimonis V, Huang T. A splice donor mutation in *NAA10* results in the dysregulation of the retinoic acid signalling pathway and causes Lenz microphthalmia syndrome. *J Med Genet* 2014;**51**:185-96 doi: 10.1136/jmedgenet-2013-101660published Online First: 2014/01/17].
8. Forrester S, Kovach MJ, Reynolds NM, Urban R, Kimonis V. Manifestations in four males with and an obligate carrier of the Lenz microphthalmia syndrome. *Am J Med Genet* 2001;**98**:92-100 Online First: 2001/06/28].
9. Graham CA, Redmond RM, Nevin NC. X-linked clinical anophthalmos. Localization of the gene to Xq27-Xq28. *Ophthalmic Paediatr Genet* 1991;**12**:43-8 Online First: 1991/03/01].
10. Lathrop GM, Lalouel JM, Julier C, Ott J. Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci U S A* 1984;**81**:3443-6 Online First: 1984/06/01].
11. Cottingham RW, Jr., Idury RM, Schaffer AA. Faster sequential genetic linkage computations. *Am J Hum Genet* 1993;**53**:252-63 Online First: 1993/07/01].

12. Schaffer AA, Gupta SK, Shriram K, Cottingham RW, Jr. Avoiding recomputation in linkage analysis. *Hum Hered* 1994;**44**:225-37 Online First: 1994/07/01].
13. Kong X, Murphy K, Raj T, He C, White PS, Matise TC. A combined linkage-physical map of the human genome. *Am J Hum Genet* 2004;**75**:1143-8 doi: 10.1086/426405published Online First: 2004/10/16].
14. Johnston JJ, Sanchez-Contreras MY, Keppler-Noreuil KM, Sapp J, Crenshaw M, Finch NA, Cormier-Daire V, Rademakers R, Sybert VP, Biesecker LG. A point mutation in *PDGFRB* causes autosomal-dominant Penttinen syndrome. *Am J Hum Genet* 2015;**97**:465-74 doi: 10.1016/j.ajhg.2015.07.009published Online First: 2015/08/19].
15. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;**536**:285-91 doi: 10.1038/nature19057published Online First: 2016/08/19].
16. Patton J, Brewer C, Chien W, Johnston JJ, Griffith AJ, Biesecker LG. A genotypic ascertainment approach to refute the association of *MYO1A* variants with non-syndromic deafness. *Eur J Hum Genet* 2016;**25**:147-49 doi: 10.1038/ejhg.2016.140published Online First: 2016/10/21].
17. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of *HpaII* and *HhaI* sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet* 1992;**51**:1229-39 Online First: 1992/12/01].
18. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15-21 doi: 10.1093/bioinformatics/bts635published Online First: 2012/10/30].
19. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 2012;**28**:2184-5 doi: 10.1093/bioinformatics/bts356published Online First: 2012/06/30].
20. Slavotinek A, Lee SS, Hamilton SP. A family with X-linked anophthalmia: exclusion of *SOX3* as a candidate gene. *Am J Med Genet A* 2005;**138A**:89-94 doi: 10.1002/ajmg.a.30872published Online First: 2005/08/23].
21. Ott J. *A Short Guide to Linkage Analysis*, 1986.

22. Cheng Y, Miura RM, Tian B. Prediction of mRNA polyadenylation sites by support vector machine. *Bioinformatics* 2006;**22**:2320-5 doi: 10.1093/bioinformatics/btl394published Online First: 2006/07/28].
23. Kalvik TV, Arnesen T. Protein N-terminal acetyltransferases in cancer. *Oncogene* 2013;**32**:269-76 doi: 10.1038/onc.2012.82published Online First: 2012/03/07].
24. Polevoda B, Arnesen T, Sherman F. A synopsis of eukaryotic Nalpha-terminal acetyltransferases: nomenclature, subunits and substrates. *BMC Proc* 2009;**3 Suppl 6**:S2 doi: 10.1186/1753-6561-3-S6-S2published Online First: 2009/08/08].
25. Myklebust LM, Van Damme P, Stove SI, Dorfel MJ, Abboud A, Kalvik TV, Grauffel C, Jonckheere V, Wu Y, Swensen J, Kaasa H, Liszczak G, Marmorstein R, Reuter N, Lyon GJ, Gevaert K, Arnesen T. Biochemical and cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects. *Hum Mol Genet* 2015;**24**:1956-76 doi: 10.1093/hmg/ddu611published Online First: 2014/12/10].
26. Lee CC, Peng SH, Shen L, Lee CF, Du TH, Kang ML, Xu GL, Upadhyay AK, Cheng X, Yan YT, Zhang Y, Juan LJ. The Role of N-alpha-acetyltransferase 10 Protein in DNA Methylation and Genomic Imprinting. *Mol Cell* 2017;**68**:89-103 e7 doi: 10.1016/j.molcel.2017.08.025published Online First: 2017/09/26].
27. Rope AF, Wang K, Evjenth R, Xing J, Johnston JJ, Swensen JJ, Johnson WE, Moore B, Huff CD, Bird LM, Carey JC, Opitz JM, Stevens CA, Jiang T, Schank C, Fain HD, Robison R, Dalley B, Chin S, South ST, Pysher TJ, Jorde LB, Hakonarson H, Lillehaug JR, Biesecker LG, Yandell M, Arnesen T, Lyon GJ. Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *Am J Hum Genet* 2011;**89**:28-43 doi: 10.1016/j.ajhg.2011.05.017published Online First: 2011/06/28].
28. Popp B, Stove SI, Endele S, Myklebust LM, Hoyer J, Sticht H, Azzarello-Burri S, Rauch A, Arnesen T, Reis A. De novo missense mutations in the NAA10 gene cause severe non-syndromic developmental delay in males and females. *Eur J Hum Genet* 2015;**23**:602-9 doi: 10.1038/ejhg.2014.150published Online First: 2014/08/08].
29. Eldomery MK, Coban-Akdemir Z, Harel T, Rosenfeld JA, Gambin T, Stray-Pedersen A, Kury S, Mercier S, Lessel D, Denecke J, Wiszniewski W, Penney S, Liu P, Bi W, Lalani SR, Schaaf CP, Wangler MF, Bacino CA, Lewis RA, Potocki L, Graham BH, Belmont JW, Scaglia F, Orange JS, Jhangiani SN, Chiang T, Doddapaneni H, Hu J, Muzny DM, Xia F, Beaudet AL, Boerwinkle E, Eng CM, Plon SE, Sutton VR, Gibbs RA, Posey JE, Yang Y, Lupski JR. Lessons learned from additional research analyses of unsolved clinical exome cases. *Genome Med* 2017;**9**:26 doi: 10.1186/s13073-017-0412-6published Online First: 2017/03/23].
30. Saunier C, Stove SI, Popp B, Gerard B, Blenski M, AhMew N, de Bie C, Goldenberg P, Isidor B, Keren B, Leheup B, Lampert L, Mignot C, Tezcan K, Mancini GM, Nava C, Wasserstein M, Bruel AL, Thevenon J, Masurel A, Duffourd Y, Kuentz P, Huet F, Riviere JB, van Slegtenhorst M, Faivre L, Piton A, Reis A, Arnesen T, Thauvin-Robinet C, Zweier C. Expanding the Phenotype Associated with NAA10-Related N-Terminal Acetylation Deficiency. *Hum Mutat* 2016;**37**:755-64 doi: 10.1002/humu.23001published Online First: 2016/04/21].

31. Cherot E, Keren B, Dubourg C, Carre W, Fradin M, Lavillaureix A, Afenjar A, Burglen L, Whalen S, Charles P, Marey I, Heide S, Jacquette A, Heron D, Doummar D, Rodriguez D, Billette de Villemeur T, Moutard ML, Guet A, Xavier J, Perisse D, Cohen D, Demurger F, Quelin C, Depienne C, Odent S, Nava C, David V, Pasquier L, Mignot C. Using medical exome sequencing to identify the causes of neurodevelopmental disorders: Experience of 2 clinical units and 216 patients. *Clin Genet* 2018;**93**:567-76 doi: 10.1111/cge.13102published Online First: 2017/07/15].
32. Thevenon J, Duffourd Y, Masurel-Paulet A, Lefebvre M, Feillet F, El Chehadeh-Djebbar S, St-Onge J, Steinmetz A, Huet F, Chouchane M, Darmency-Stamboul V, Callier P, Thauvin-Robinet C, Faivre L, Riviere JB. Diagnostic odyssey in severe neurodevelopmental disorders: toward clinical whole-exome sequencing as a first-line diagnostic test. *Clin Genet* 2016;**89**:700-7 doi: 10.1111/cge.12732published Online First: 2016/01/13].
33. Zhang H, Hu J, Recce M, Tian B. PolyA_DB: a database for mammalian mRNA polyadenylation. *Nucleic Acids Res* 2005;**33**:D116-20 doi: 10.1093/nar/gki055published Online First: 2004/12/21].
34. Sheets MD, Ogg SC, Wickens MP. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res* 1990;**18**:5799-805 Online First: 1990/10/11].
35. Jankovic L, Efremov GD, Petkov G, Kattamis C, George E, Yang KG, Stoming TA, Huisman TH. Two novel polyadenylation mutations leading to beta(+)-thalassemia. *Br J Haematol* 1990;**75**:122-6 Online First: 1990/05/01].
36. Bilgen T, Clark OA, Ozturk Z, Akif Yesilipek M, Keser I. Two novel mutations in the 3' untranslated region of the beta-globin gene that are associated with the mild phenotype of beta thalassemia. *Int J Lab Hematol* 2013;**35**:26-30 doi: 10.1111/j.1751-553X.2012.01456.xpublished Online First: 2012/08/07].
37. Gieselmann V, Polten A, Kreysing J, von Figura K. Arylsulfatase A pseudodeficiency: loss of a polyadenylylation signal and N-glycosylation site. *Proc Natl Acad Sci U S A* 1989;**86**:9436-40 Online First: 1989/12/01].
38. Gyawali S, Subaran R, Weissman MM, Hershkowitz D, McKenna MC, Talati A, Fyer AJ, Wickramaratne P, Adams PB, Hodge SE, Schmidt CJ, Bannon MJ, Glatt CE. Association of a polyadenylation polymorphism in the serotonin transporter and panic disorder. *Biol Psychiatry* 2010;**67**:331-8 doi: 10.1016/j.biopsych.2009.10.015published Online First: 2009/12/09].
39. Colgan DF, Manley JL. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* 1997;**11**:2755-66 Online First: 1997/11/14].
40. Shi Y, Manley JL. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev* 2015;**29**:889-97 doi: 10.1101/gad.261974.115published Online First: 2015/05/03].
41. Sun Y, Zhang Y, Hamilton K, Manley JL, Shi Y, Walz T, Tong L. Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proc Natl Acad Sci U S A* 2018;**115**:E1419-E28 doi: 10.1073/pnas.1718723115published Online First: 2017/12/07].

42. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 2014;**133**:1-9 doi: 10.1007/s00439-013-1358-4published Online First: 2013/10/01].
43. Giordano PC, Bouva MJ, Van Delft P, Akkerman N, Kappers-Klunne MC, Harteveld CL. A new polyadenylation site mutation associated with a mild beta-thalassemia phenotype. *Haematologica* 2005;**90**:551-2 Online First: 2005/04/12].
44. Berkovits BD, Mayr C. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* 2015;**522**:363-7 doi: 10.1038/nature14321published Online First: 2015/04/22].
45. Mayr C. Regulation by 3'-Untranslated Regions. *Annu Rev Genet* 2017;**51**:171-94 doi: 10.1146/annurev-genet-120116-024704published Online First: 2017/08/31].
46. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee ALQA. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;**17**:405-24 doi: 10.1038/gim.2015.30published Online First: 2015/03/06].

Figure 1. Pedigrees for families 1 to 3. Clinically affected individuals are depicted by filled symbols, black symbols depict individuals with eye findings, grey symbols depict individuals without eye findings, genotypes of tested individuals are noted.

Figure 2. Reverse transcription and quantitative PCR analysis of *NAA10* expression in mRNA, data were normalized to *ACTB* mRNA levels. Panel A, relative *NAA10* expression in whole blood from affected individuals (family 1, individuals III-2, III-5 and IV-3), carrier females (family 1, individuals II-2, II-6 and III-15) and male and female control individuals (C1-C4). All values are shown relative to control C1. Panel B, *NAA10* expression levels in lymphoblasts from affected male IV-6 shown relative to the expression level in unaffected male V-9, family 2.

Figure 3. RNA-seq data comparing reads from four affected individuals, two carrier females and three control individuals, one female and two males, for *NAA10*. The variant positions in the *NAA10* polyadenylation signal are marked by a filled red arrow. Transcription in affected individuals (family 1, individuals III-2, III-5 and IV-3 and family 2, individual IV-6) and carrier females (family 1, individuals II-2 and III-15) continues past the normal polyadenylation cleavage site and uses a cryptic signal approximately 600 bp downstream at an alternate polyadenylation cleavage site depicted by an open red arrow. *NAA10* gene models are from GENCODE Version 19 as included in the UCSC genome browser (GRCh37/hg19).

Figure 4. Comparison of the effects of AAUAAA mutations on cleavage and poly (A) addition. Variants identified in this study are noted with an asterisk. Figure is reproduced from Sheets et al. 1990 with permission from Oxford University Press.[34]

Table 1. *NAA10* polyadenylation signal variants identified in microphthalmia/anophthalmia and review of polyadenylation signal alterations as reported in HGMD.

Gene	**WT[a]**	**Variant**	**HGMD ID**	**HGMD pathogenicity**	**ClinVar Descriptor**	**GRCh37 Position & Variation**	**ClinVar variation ID**	**ClinVar Assertion**	**PMID**
NAA10		AATAGA				ChrX:153195397T>C			
NAA10		GATAAA				ChrX:153195401T>C			
NAA10		AGTAAA				ChrX:153195400T>C			
BMP1	AGTAAA	AGCAAA	CR150372	DM	NM_001199.3(BMP1):c.*241T>C	Chr8:22058957T>C	190231	Pathogenic	25214535
F9		AATGAA	CR005437	DM					11013449
FOXP3		AATGAA	CR014834	DM					11685453
FOXP3		AATAAG	CR097218	DM					19471859
GLA	ATTAAAga	ATTAGA	CD031841	DM					12796853
HBA2		AATAAG	CR830007	DM					6646217
HBA2		AATGAA	CR920785	DM	NM_000517.4(HBA2):c.*92A>G	Chr16:223691A>G	15647	Pathogenic/Likely Pathogenic	1581238
HBA2		AATAAC	CR106042	DM					19912309
HBA2	AATAAAgt	AATAGT	CD941949	DM					7803252
HBB		AAAAAA	CR045224	DM					15481893
HBB		AAGAAA	CR014260	DM					11300343
HBB		AATATA	CR057232	DM					15820953
HBB		AACAAA	CR850010	DM	NM_000518.4(HBB):c.*110T>C	Chr11:5246718A>G	36332	Pathogenic	4018033
HBB		AATAAG	CR880076	DM	NM_000518.4(HBB):c.*113A>G	Chr11:5246715T>C	15473	Pathogenic	3048433

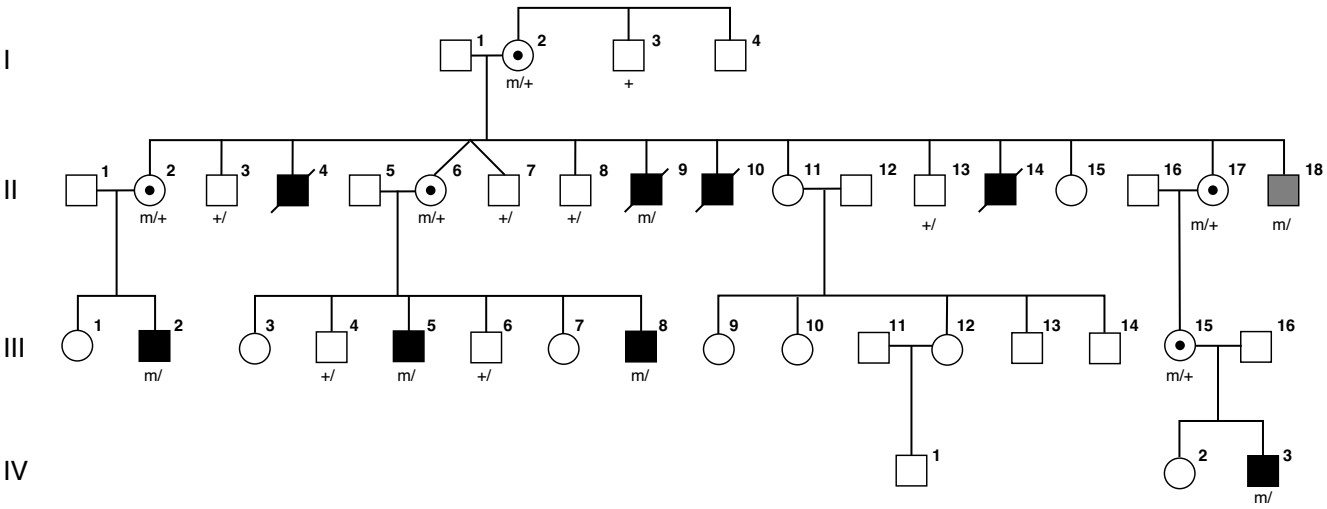
HBB		AATGAA	CR900265	DM	NM_000518.4(HBB):c.*111A>G	Chr11:5246717T>C	15488	Pathogenic	2375910
HBB		AATAGA	CR900266	DM	NM_000518.4(HBB):c.*112A>G	Chr11:5246716T>C	15476	Pathogenic	2375910
HBB		CATAAA	CR016252	DM					11722440
HBB		GATAAA	CR127145	DM					22862814
HBD		AATTAA	CR109506	DM					20854114
IGF1	AATATA	AAAATA	CR033689	DM					14684690
IL2RG		AATAAG	CR0910465	DM					19841577
INS		AATAAG	CR101141	DM	NM_000207.2(INS):c.*59A>G	Chr11:2181023T>C	65581	Pathogenic	20133622
ITGA2B		AACAAA	CR153724	DM					25728920
ABCG2	AAATGA	AAGTGA	CR1718077	DP					28930109
ARSA	AATAAC	AGTAAC	CR890137	DP	NM_000487.5(ARSA):c.*96A>G	Chr22:51063477T>C	3049	Conflicting	2574462
HES7	AAAATC	AAGATC	CR1510755	DM?					25928698
NAT1		AAAAAA	CR1111021	FP					7585580
PEX6	AATAAAcaca	AACACA	CD1716837	FP					29220678
PTGFR		AATGAA	CR156013	DFP					25977569
SLC6A4	ATTAAC	AGTAAC	CR102248	DFP					19969287
TP53		AATACA	CR118782	DFP					21946351

[a]Wild type is AATAAA unless otherwise specified. Underlined nucleotides are deleted and the nucleotides given in lower case are shifted into the variant hexamer. Thymidine is specified instead of uracil, referencing the genomic sequence, rather than the mRNA.

Figure 1

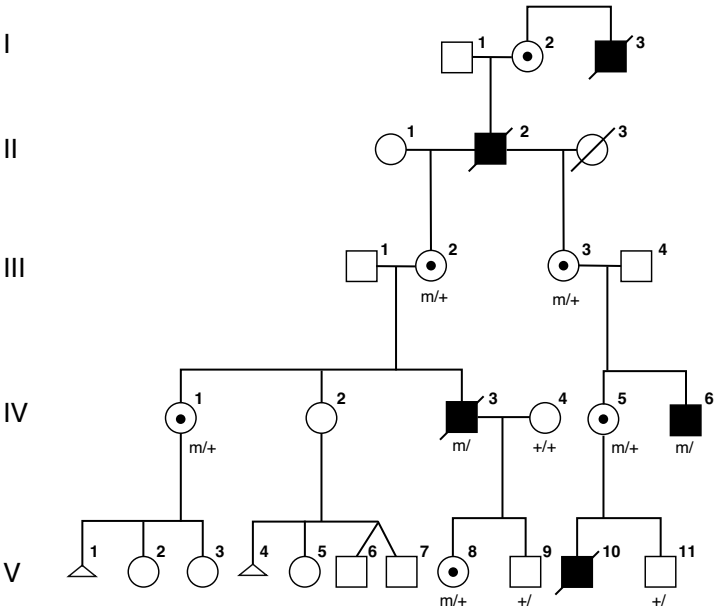
A

c.*43A>G



B

c.*39A>G



C

c.*40A>G

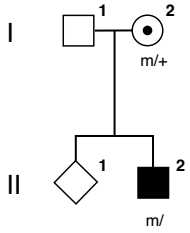


Figure 2

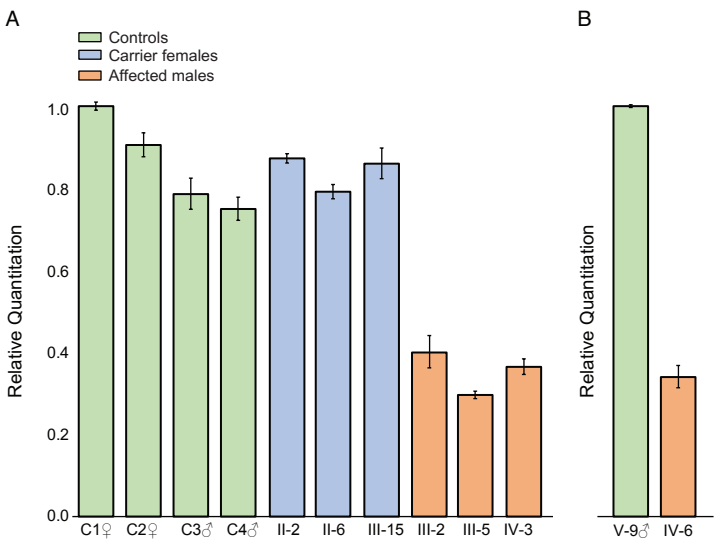

A
Controls
Carrier females
Affected males
Relative Quantitation
1.0
0.8
0.6
0.4
0.2
0.0
C1♀
C2♀
C3♂
C4♂
II-2
II-6
III-15
III-2
III-5
IV-3
B
Relative Quantitation
V-9♂
IV-6

Figure 3

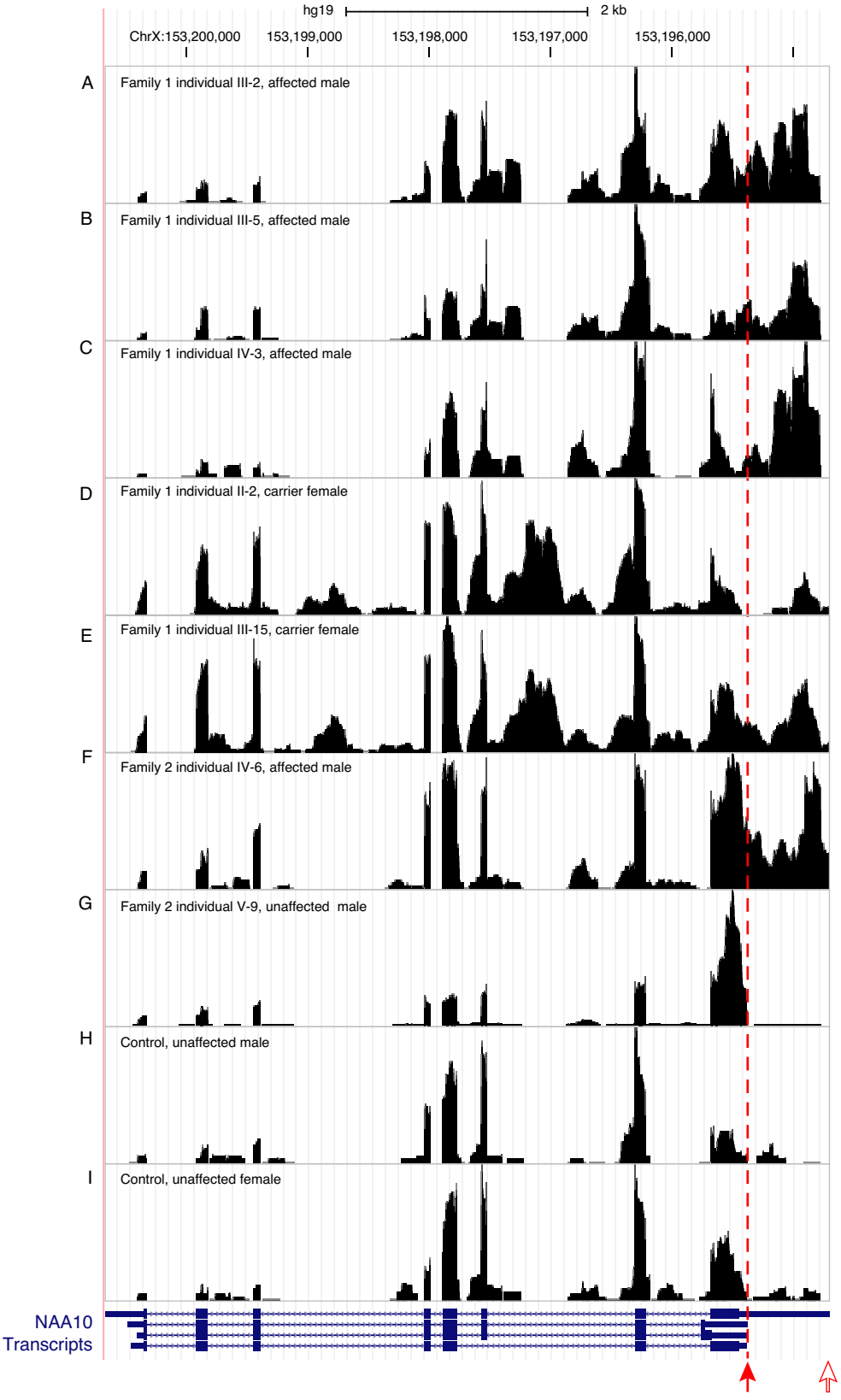
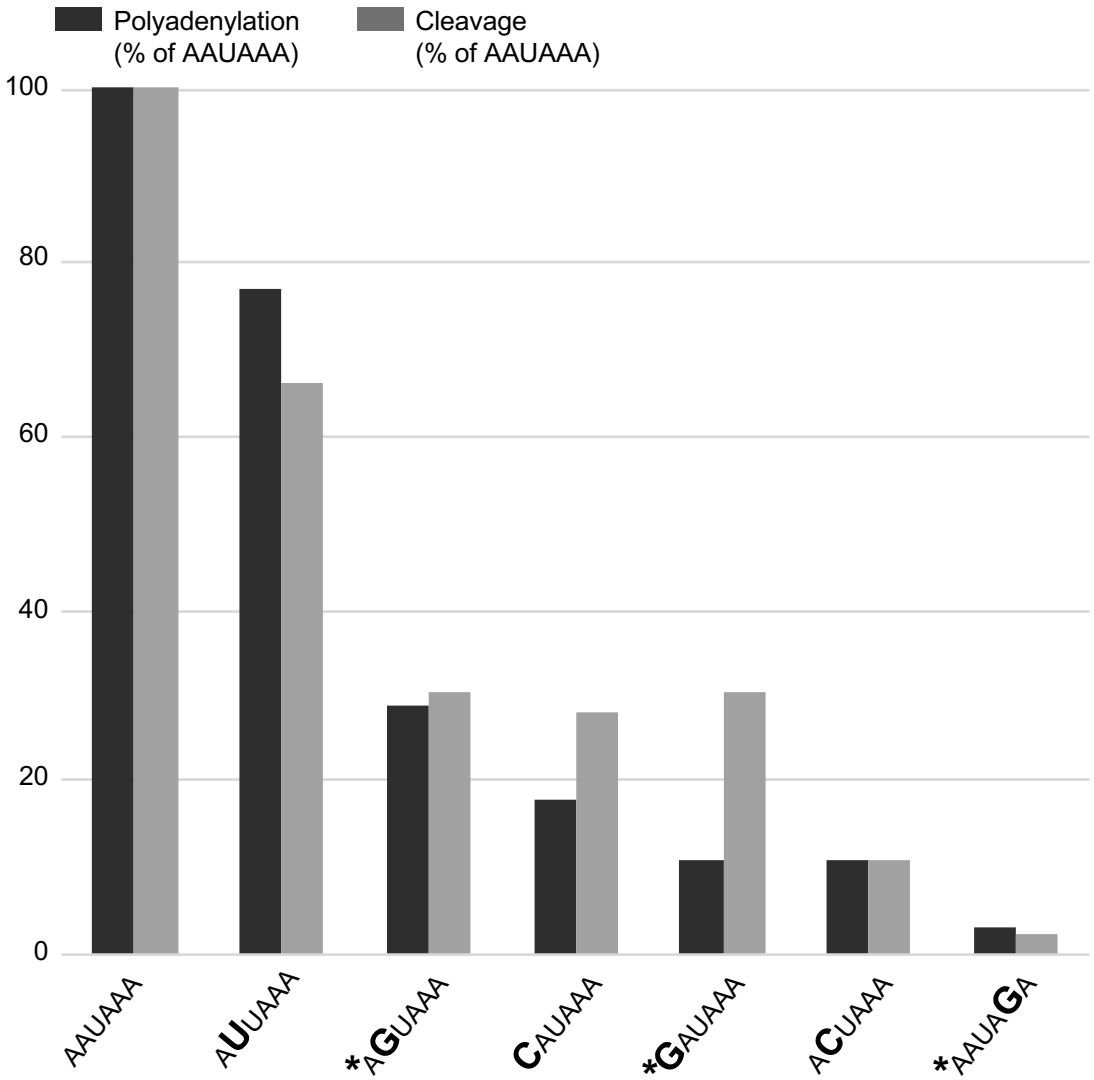

hg19
2 kb
ChrX:153,200,000
153,199,000
153,198,000
153,197,000
153,196,000
A
Family 1 individual III-2, affected male
B
Family 1 individual III-5, affected male
C
Family 1 individual IV-3, affected male
D
Family 1 individual II-2, carrier female
E
Family 1 individual III-15, carrier female
F
Family 2 individual IV-6, affected male
G
Family 2 individual V-9, unaffected male
H
Control, unaffected male
I
Control, unaffected female
NAA10
Transcripts

Figure 4


Polyadenylation
(% of AAUAAA)
Cleavage
(% of AAUAAA)
100
80
60
40
20
0
AAUAAA
AUUAAA
*AGUAAA
CAUAAA
*GAUAAA
ACUAAA
*AAUAGA